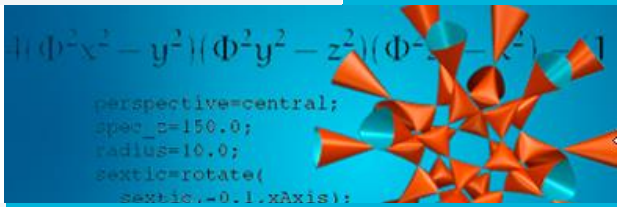
Philipps Universität Marburg

Towards an EBV Analyzer based on VAT

Christian Beilschmidt[1], Johannes Drönner[1], Néstor Fernández[2], Christian Langer[2], Michael Mattig[1], Bernhard Seeger[1]

[1]*University of Marburg, Germany*
[2]*German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany*

Corresponding author(s) e-mail: {beilschmidt, droenner}@mathematik.uni-marburg.de,



Motivation

- EBVs are relevant for global and local questions

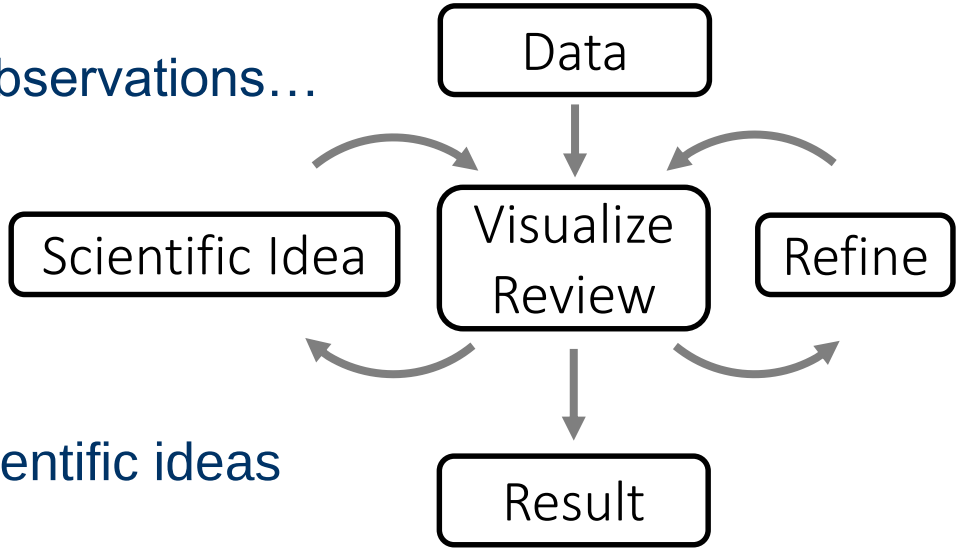
  → **Accessibility?**

- VAT allows domain experts to explore data & create analytic workflows
  - Reuse of existing functionality / avoid re-implementation
  - Privacy preserving / combine EBVs easily with non-public data
  - Support of provenance & lineage (reproduce)

→Interactive reports: views for end-users & decision makers

Concepts: Data-Driven Research

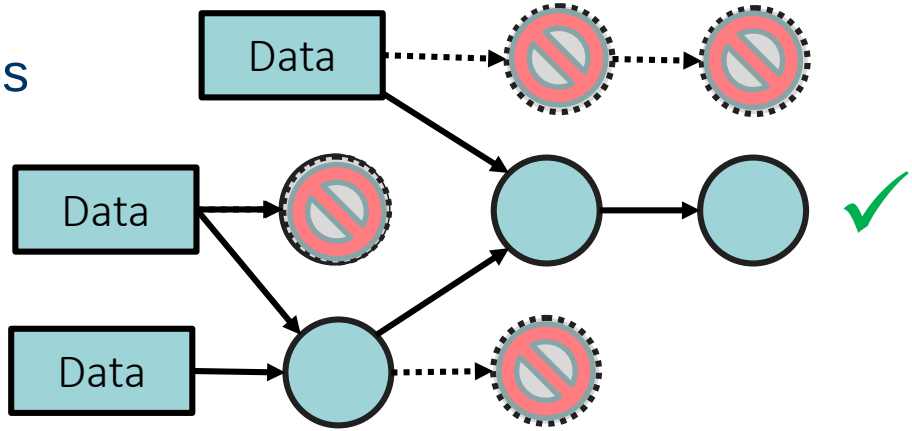
- Data availability
  - Satellites, sensors and observations…



- Visualization triggers new scientific ideas

Concepts: Explorative Workflows

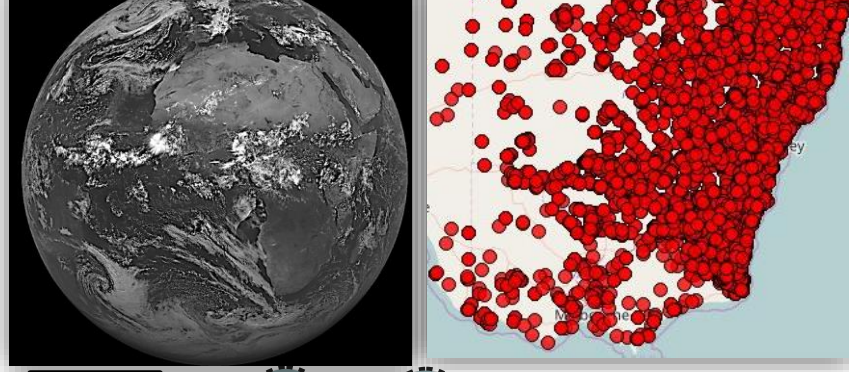
- Explore data and processing steps
  - Multiple paths

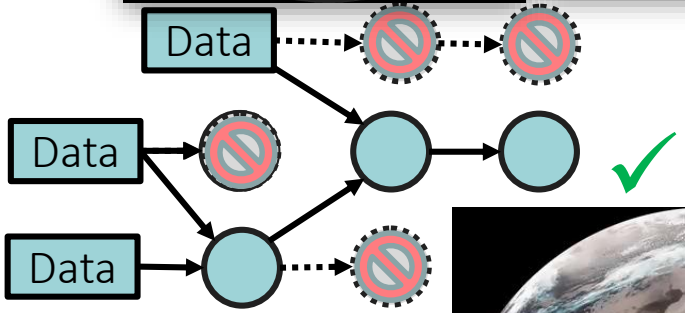


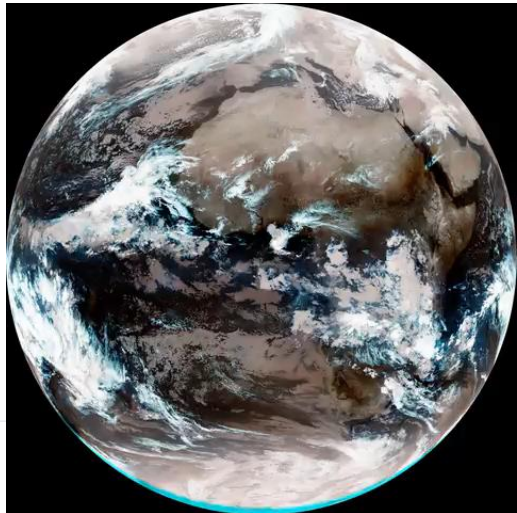
- Citations + reproducibility
- → Enables application builder for interactive reports
  - Views for end-users & decision makers

Challenges

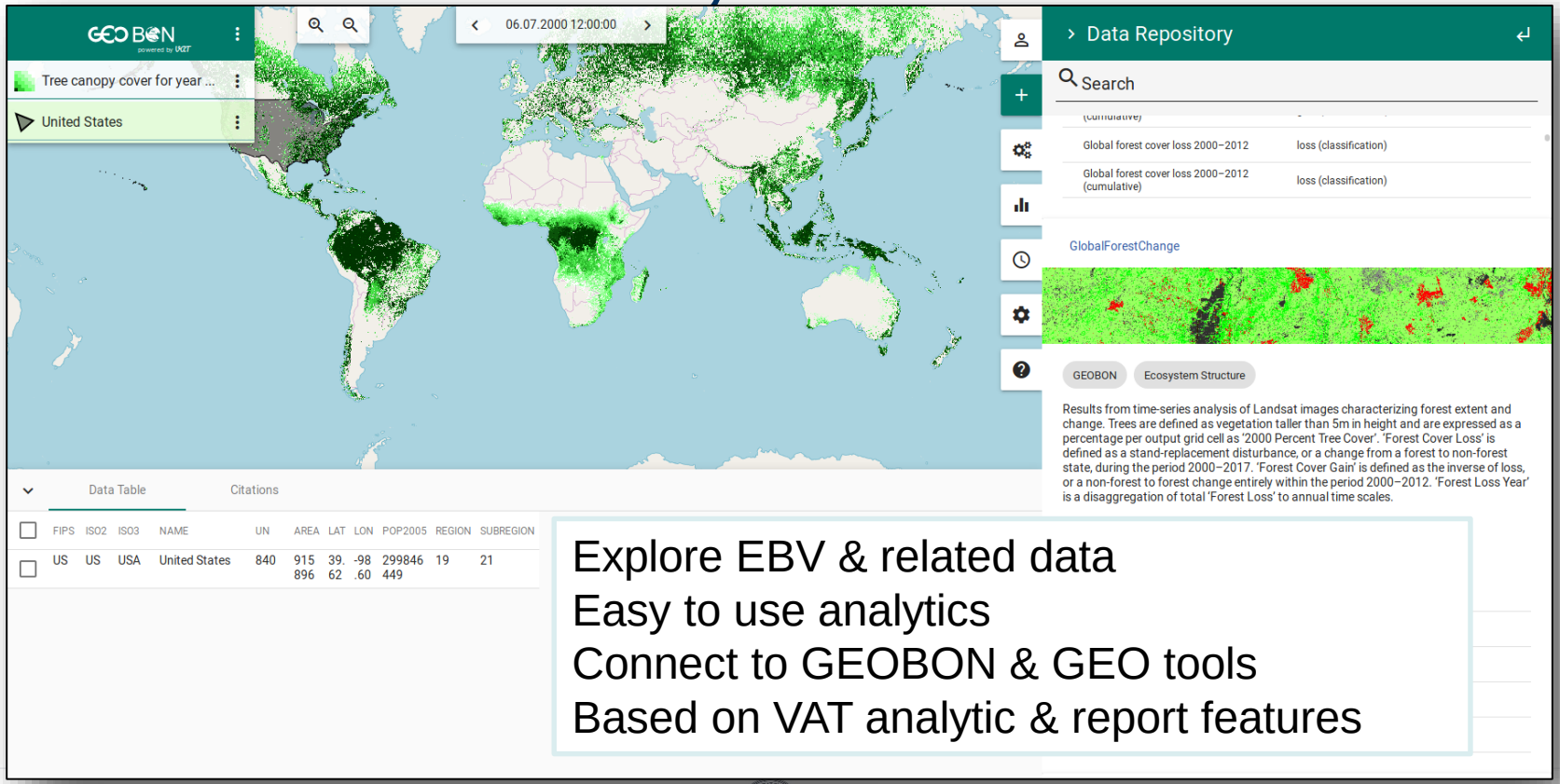
- Visualization
  - Big / heterogeneous data
- Explorative workflows
  - Data lineage / citations
- Time as an integral dimension
  - Track changes over time
  - Detect temporal patterns







EBV Browser and Analyzer



Explore EBV & related data
Easy to use analytics
Connect to GEOBON & GEO tools
Based on VAT analytic & report features

Use-case: Forest loss in northern America

Global Forest Change

(Organization: University of Maryland; Published by Hansen, Potapov, Moore, Hancher et al.)

Results from time-series analysis of Landsat images characterizing forest extent and change.

- Temporal development of loss (2000 – 2012)
- User-defined area of interest
  - United States or polygon

- Generic use-case → also interesting for other EBVs & questions!

GEO BON
powered by UFZ

United States

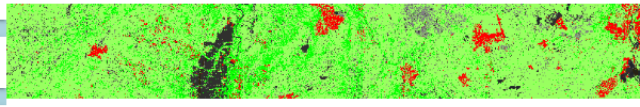
06.07.2000 12:00:00

Data Repository

Search

Global forest cover loss 2000–2012	loss (classification)
Global forest cover loss 2000–2012 (cumulative)	loss (classification)

GlobalForestChange



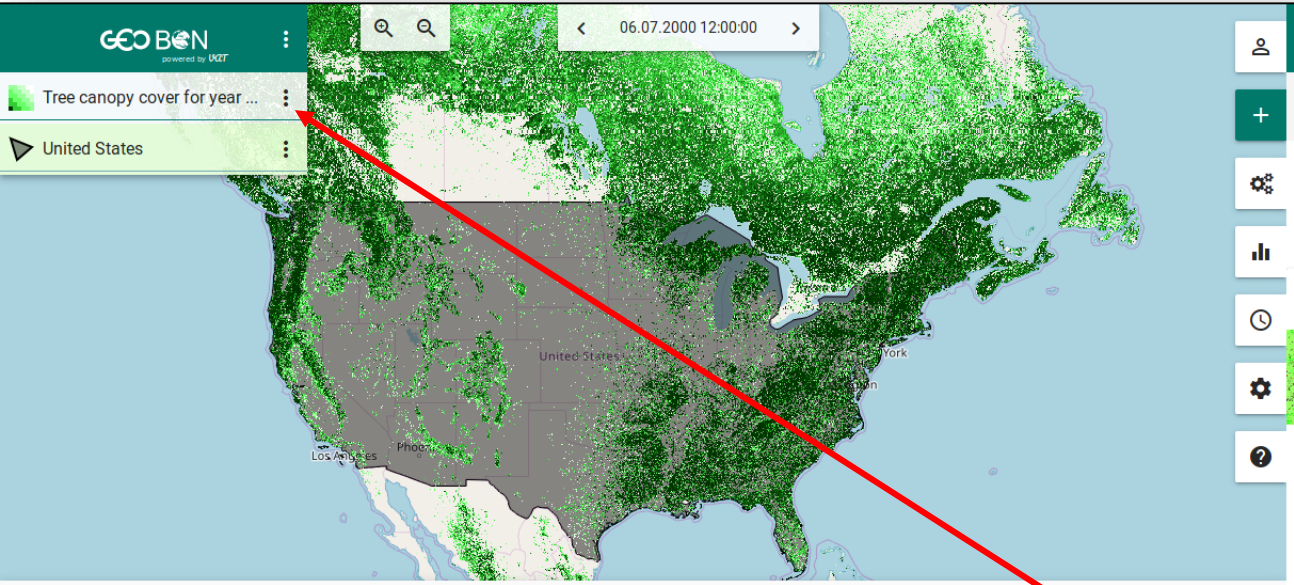
GEOBON | Ecosystem Structure

Results from time-series analysis of Landsat images characterizing forest extent and change. Trees are defined as vegetation taller than 5m in height and are expressed as a percentage per output grid cell as '2000 Percent Tree Cover'. 'Forest Cover Loss' is defined as a stand-replacement disturbance, or a change from a forest to non-forest state, during the period 2000–2017. 'Forest Cover Gain' is defined as the inverse of loss, or a non-forest to forest change entirely within the period 2000–2012. 'Forest Loss Year' is a disaggregation of total 'Forest Loss' to annual time scales.

More information: Website

Channel	Measurement
Tree canopy cover for year 2000	tree cover (%)
Global forest cover gain 2000–2012	gain (classification)
Global forest cover loss 2000–2012	loss (classification)
Year of gross forest cover loss event	loss year (unknown)

Data Table | Citations



Data Table | Citations

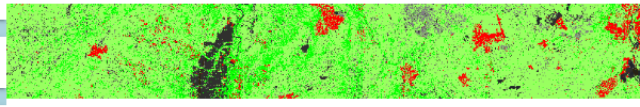
	FIPS	ISO2	ISO3	NAME	UN	AREA	LAT	LON	POP2005	REGION	SUBREGION
☐	US	US	USA	United States	840	915 896	39. 62	-98 .60	299846 449	19	21

Data Repository

Search

Global forest cover loss 2000–2012	loss (classification)
Global forest cover loss 2000–2012 (cumulative)	loss (classification)

GlobalForestChange

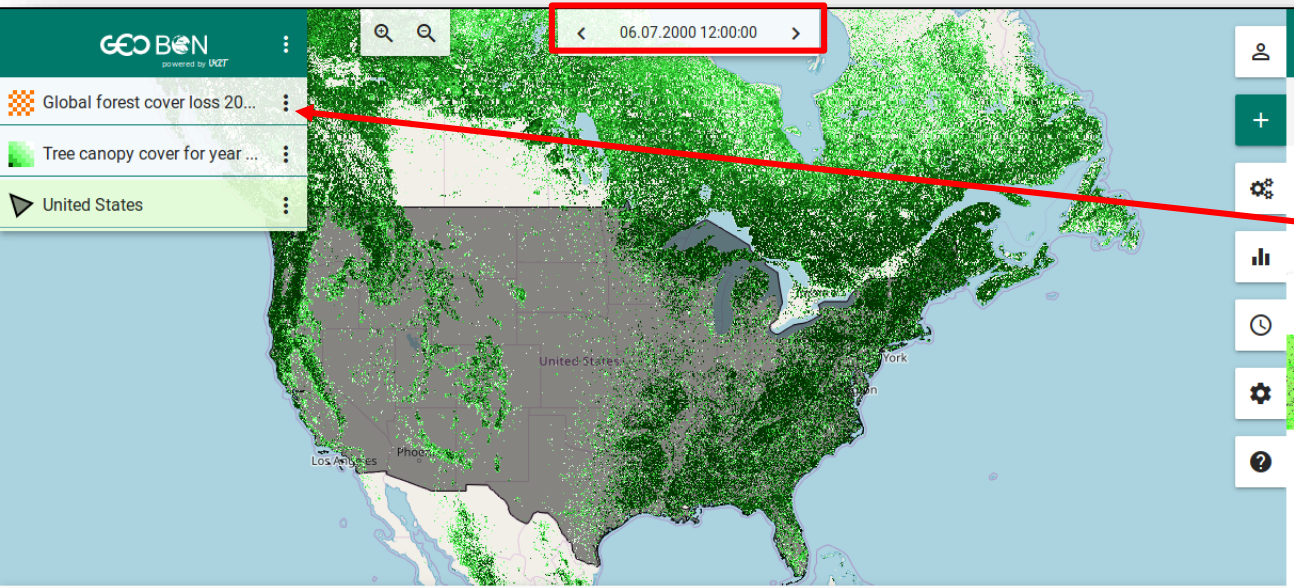


GEOBON | Ecosystem Structure

Results from time-series analysis of Landsat images characterizing forest extent and change. Trees are defined as vegetation taller than 5m in height and are expressed as a percentage per output grid cell as '2000 Percent Tree Cover'. 'Forest Cover Loss' is defined as a stand-replacement disturbance, or a change from a forest to non-forest state, during the period 2000–2017. 'Forest Cover Gain' is defined as the inverse of loss, or a non-forest to forest change entirely within the period 2000–2012. 'Forest Loss Year' is a disaggregation of total 'Forest Loss' to annual time scales.

More information: Website

Channel	Measurement
Tree canopy cover for year 2000	tree cover (%)
Global forest cover gain 2000–2012	gain (classification)
Global forest cover loss 2000–2012	loss (classification)
Year of gross forest cover loss event	loss year (unknown)



Data Table | Citations

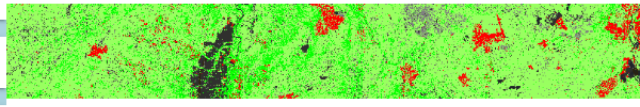
FIPS	ISO2	ISO3	NAME	UN	AREA	LAT	LON	POP2005	REGION	SUBREGION
US	US	USA	United States	840	915 896	39. 62	-98 .60	299846 449	19	21

Data Repository

Search

Global forest cover loss 2000–2012	loss (classification)
Global forest cover loss 2000–2012 (cumulative)	loss (classification)

GlobalForestChange

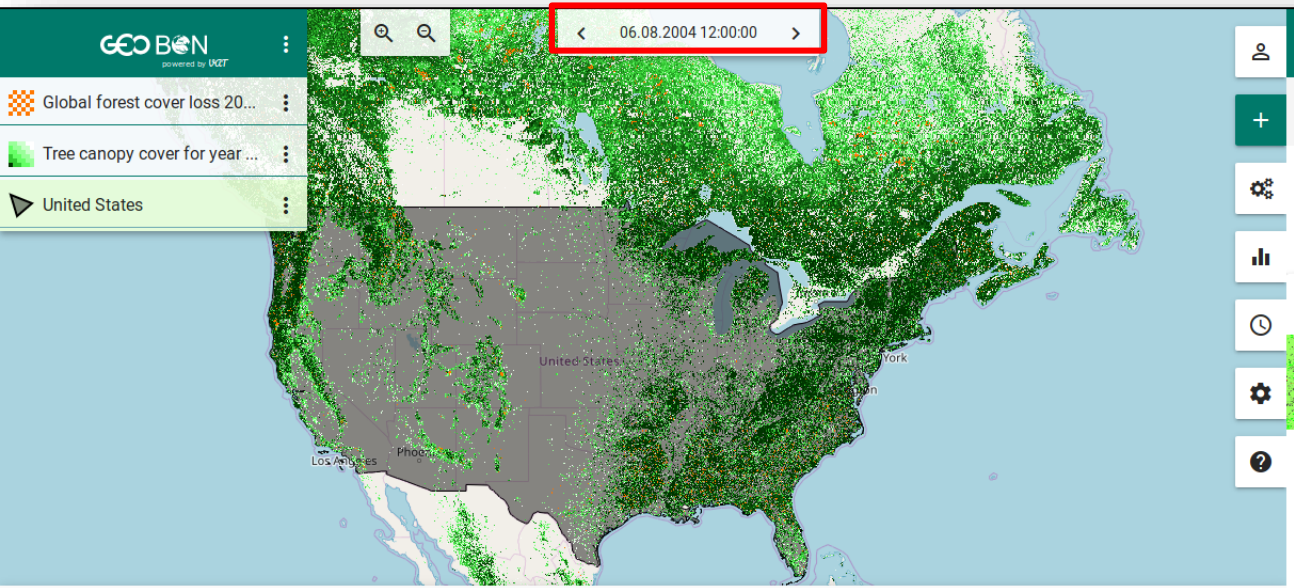


GEOBON | Ecosystem Structure

Results from time-series analysis of Landsat images characterizing forest extent and change. Trees are defined as vegetation taller than 5m in height and are expressed as a percentage per output grid cell as '2000 Percent Tree Cover'. 'Forest Cover Loss' is defined as a stand-replacement disturbance, or a change from a forest to non-forest state, during the period 2000–2017. 'Forest Cover Gain' is defined as the inverse of loss, or a non-forest to forest change entirely within the period 2000–2012. 'Forest Loss Year' is a disaggregation of total 'Forest Loss' to annual time scales.

More information: Website

Channel	Measurement
Tree canopy cover for year 2000	tree cover (%)
Global forest cover gain 2000–2012	gain (classification)
Global forest cover loss 2000–2012	loss (classification)
Year of gross forest cover loss event	loss year (unknown)



Data Table | Citations

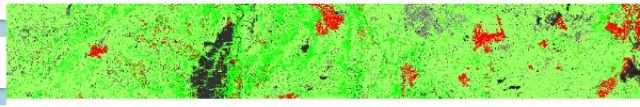
	FIPS	ISO2	ISO3	NAME	UN	AREA	LAT	LON	POP2005	REGION	SUBREGION
☐	US	US	USA	United States	840	915 896	39. 62	-98 .60	299846 449	19	21

Data Repository

Search

Global forest cover loss 2000–2012	loss (classification)
Global forest cover loss 2000–2012 (cumulative)	loss (classification)

GlobalForestChange

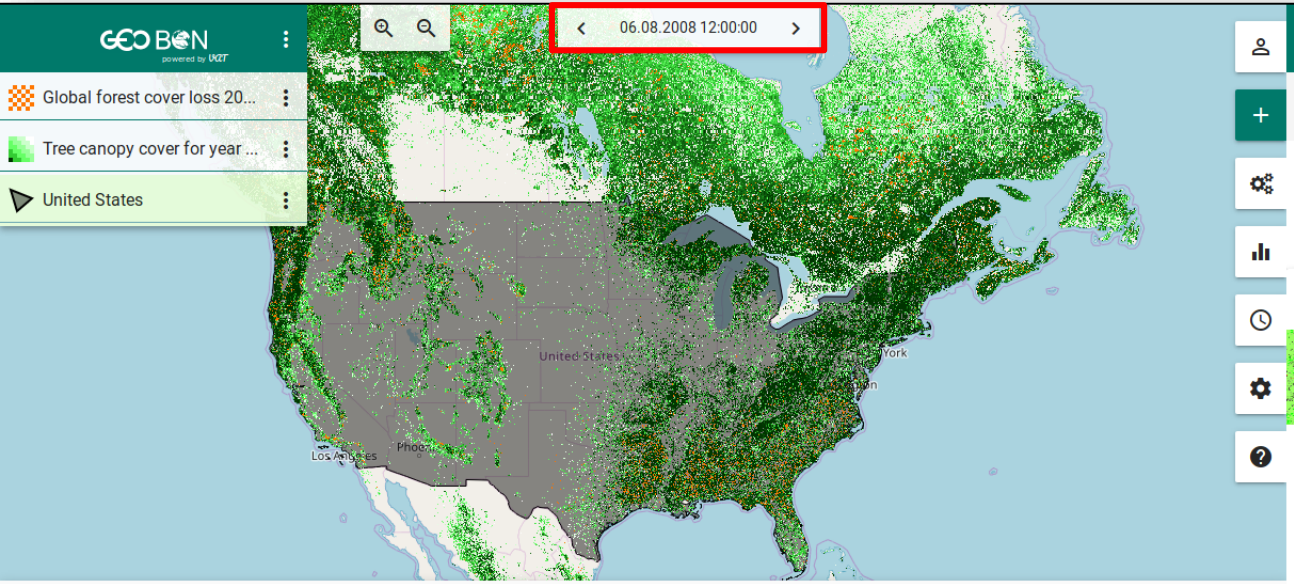


GEOBON | Ecosystem Structure

Results from time-series analysis of Landsat images characterizing forest extent and change. Trees are defined as vegetation taller than 5m in height and are expressed as a percentage per output grid cell as '2000 Percent Tree Cover'. 'Forest Cover Loss' is defined as a stand-replacement disturbance, or a change from a forest to non-forest state, during the period 2000–2017. 'Forest Cover Gain' is defined as the inverse of loss, or a non-forest to forest change entirely within the period 2000–2012. 'Forest Loss Year' is a disaggregation of total 'Forest Loss' to annual time scales.

More information: Website

Channel	Measurement
Tree canopy cover for year 2000	tree cover (%)
Global forest cover gain 2000–2012	gain (classification)
Global forest cover loss 2000–2012	loss (classification)
Year of gross forest cover loss event	loss year (unknown)



Data Table | Citations

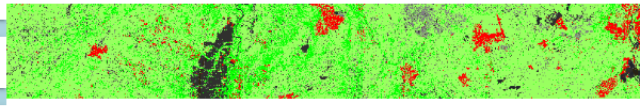
FIPS	ISO2	ISO3	NAME	UN	AREA	LAT	LON	POP2005	REGION	SUBREGION
US	US	USA	United States	840	915 896	39. 62	-98 .60	299846 449	19	21

Data Repository

Search

Global forest cover loss 2000–2012	loss (classification)
Global forest cover loss 2000–2012 (cumulative)	loss (classification)

GlobalForestChange



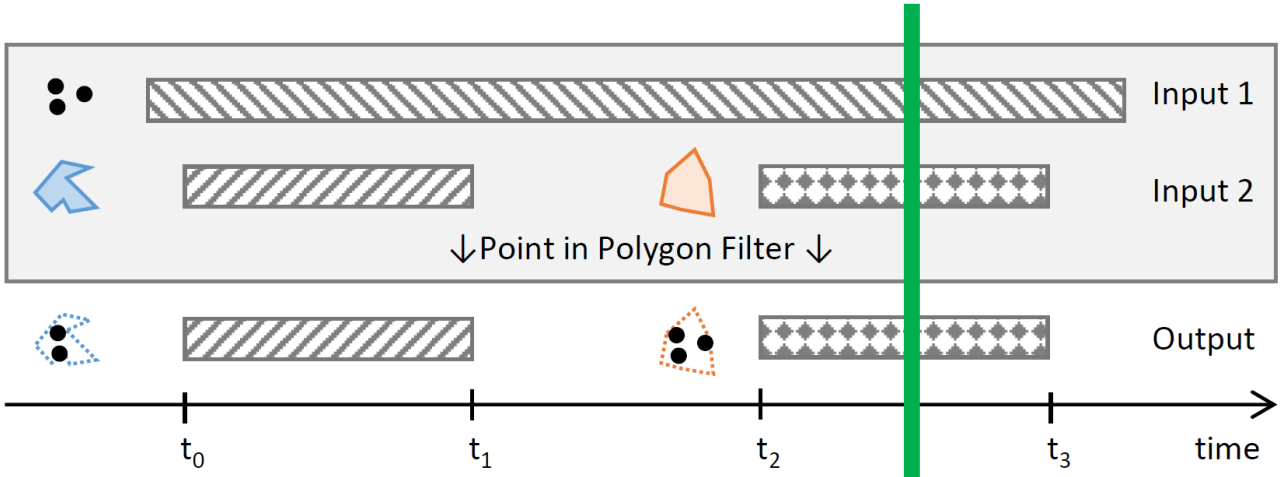
GEOBON | Ecosystem Structure

Results from time-series analysis of Landsat images characterizing forest extent and change. Trees are defined as vegetation taller than 5m in height and are expressed as a percentage per output grid cell as '2000 Percent Tree Cover'. 'Forest Cover Loss' is defined as a stand-replacement disturbance, or a change from a forest to non-forest state, during the period 2000–2017. 'Forest Cover Gain' is defined as the inverse of loss, or a non-forest to forest change entirely within the period 2000–2012. 'Forest Loss Year' is a disaggregation of total 'Forest Loss' to annual time scales.

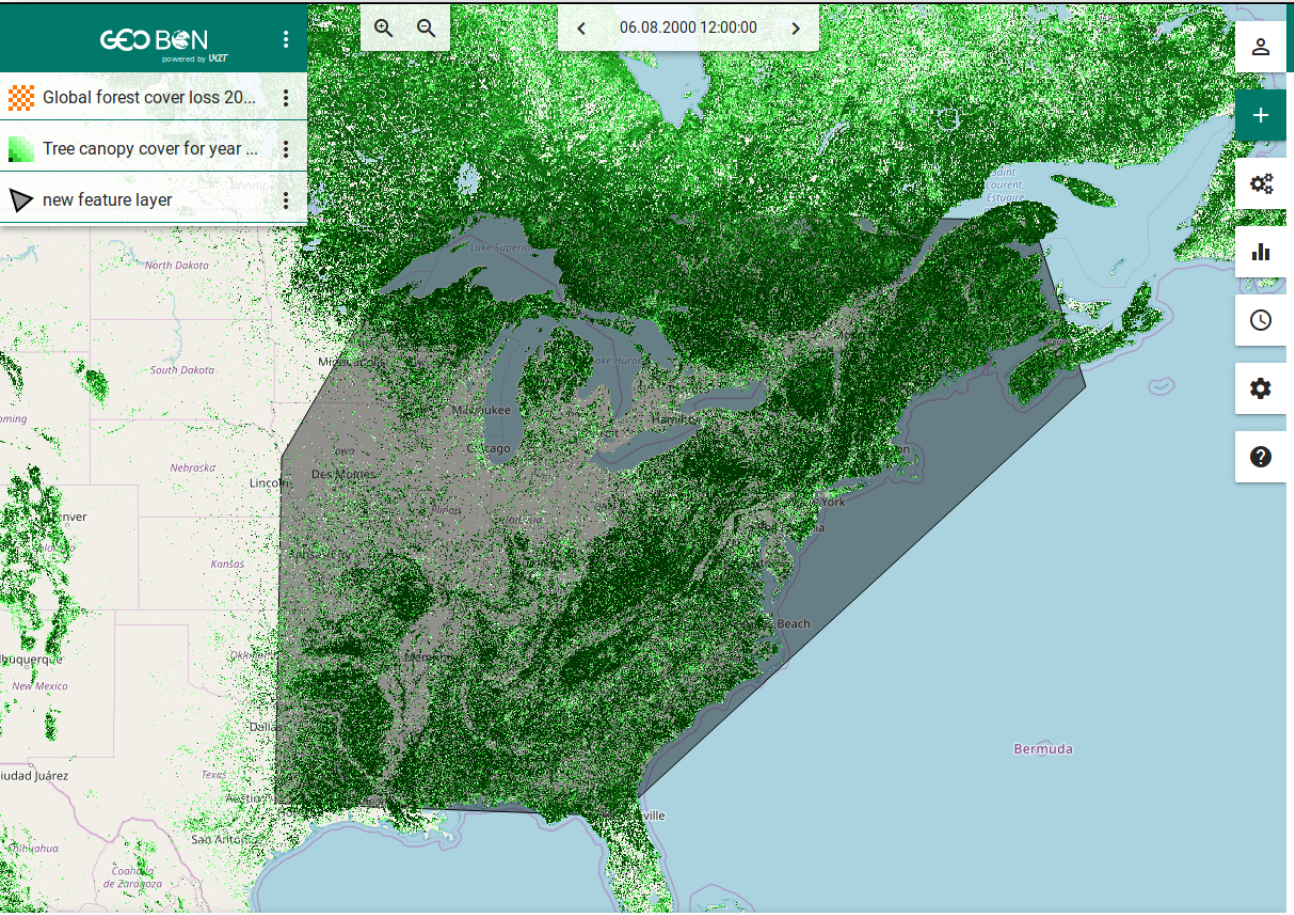
More information: Website

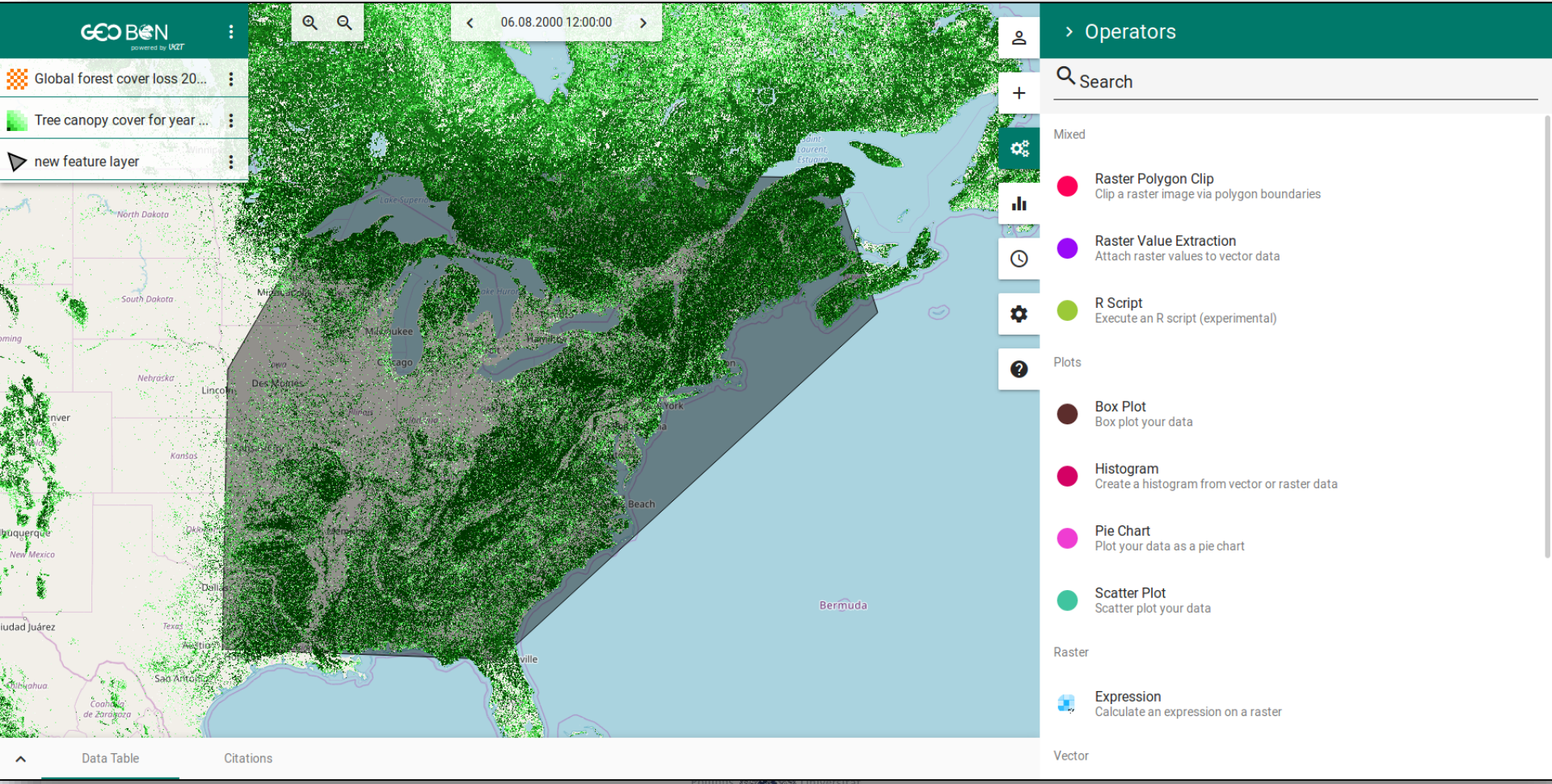
Channel	Measurement
Tree canopy cover for year 2000	tree cover (%)
Global forest cover gain 2000–2012	gain (classification)
Global forest cover loss 2000–2012	loss (classification)
Year of gross forest cover loss event	loss year (unknown)

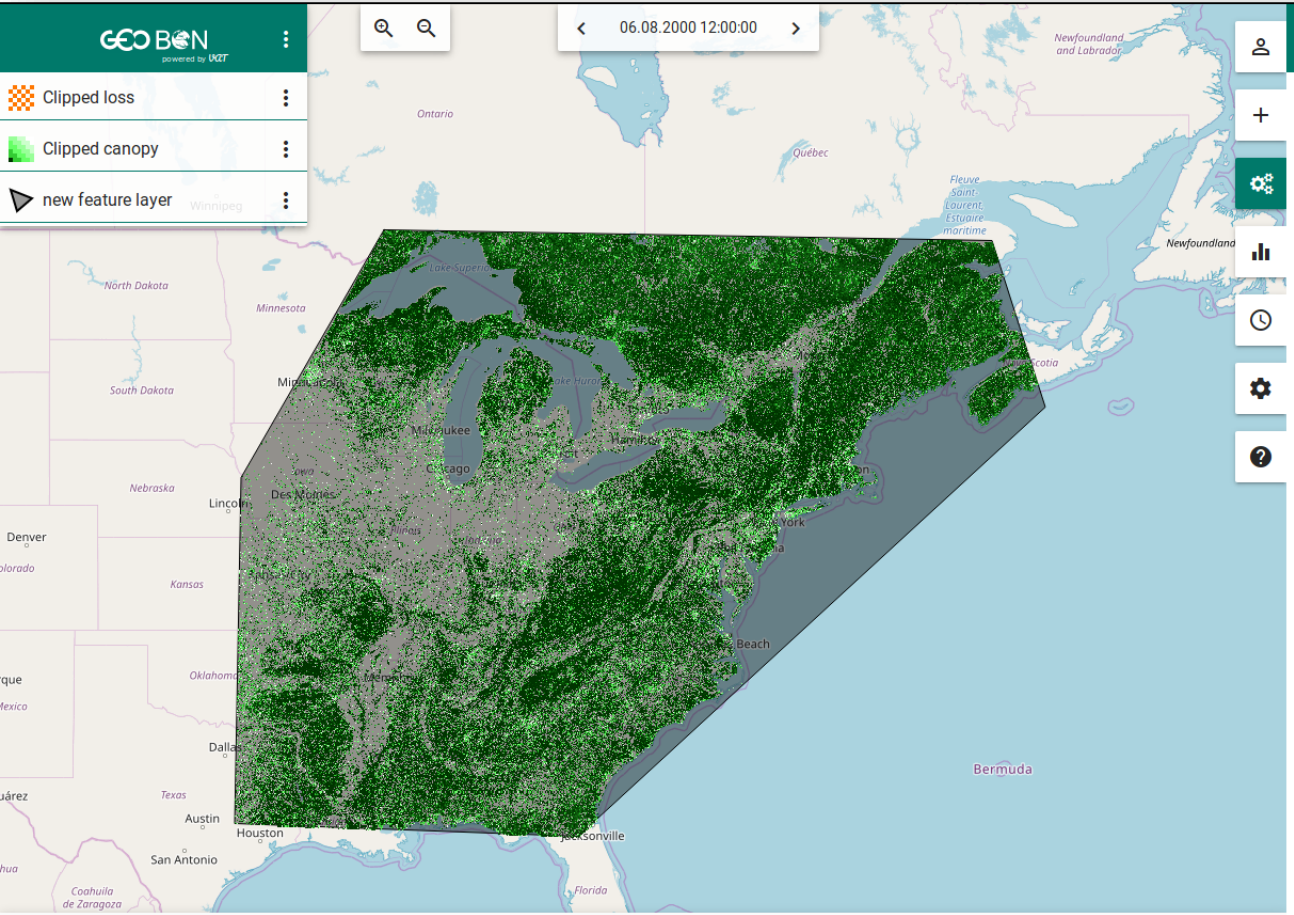
Time as Integral Dimension



- All datasets are time-series
  - Individual temporal validity
- Operations create new time-series


GEO BON
powered by UZH
Global forest cover loss 20...
Tree canopy cover for year ...
new feature layer
06.08.2000 12:00:00
North Dakota
South Dakota
Nebraska
Kansas
New Mexico
Texas
Milwaukee
Chicago
Des Moines
Lincoln
Denver
Albuquerque
Ciudad Juárez
San Antonio
Bermuda
Beach
Draw Features
Feature type
Polygon
Start
End
Cancel
Data Table
Citations


GEO BON
powered by UZH
Global forest cover loss 20...
Tree canopy cover for year ...
new feature layer
06.08.2000 12:00:00
Operators
Search
Mixed
Raster Polygon Clip
Clip a raster image via polygon boundaries
Raster Value Extraction
Attach raster values to vector data
R Script
Execute an R script (experimental)
Plots
Box Plot
Box plot your data
Histogram
Create a histogram from vector or raster data
Pie Chart
Plot your data as a pie chart
Scatter Plot
Scatter plot your data
Raster
Expression
Calculate an expression on a raster
Vector
Data Table
Citations
North Dakota
South Dakota
Nebraska
Kansas
New Mexico
Texas
Milwaukee
Chicago
Des Moines
Bermuda



Calculate Expression on Raster

Raster

Select input Raster

Input A

Clipped loss

Input B

Clipped canopy

Configuration

Specify the operator

Use A to reference the existing pixel of the first raster, B for the second one, etc.

Expression

(A==1) ? 0 : B

Output Data Type

Byte (like layers A,B)

Minimum Value

0

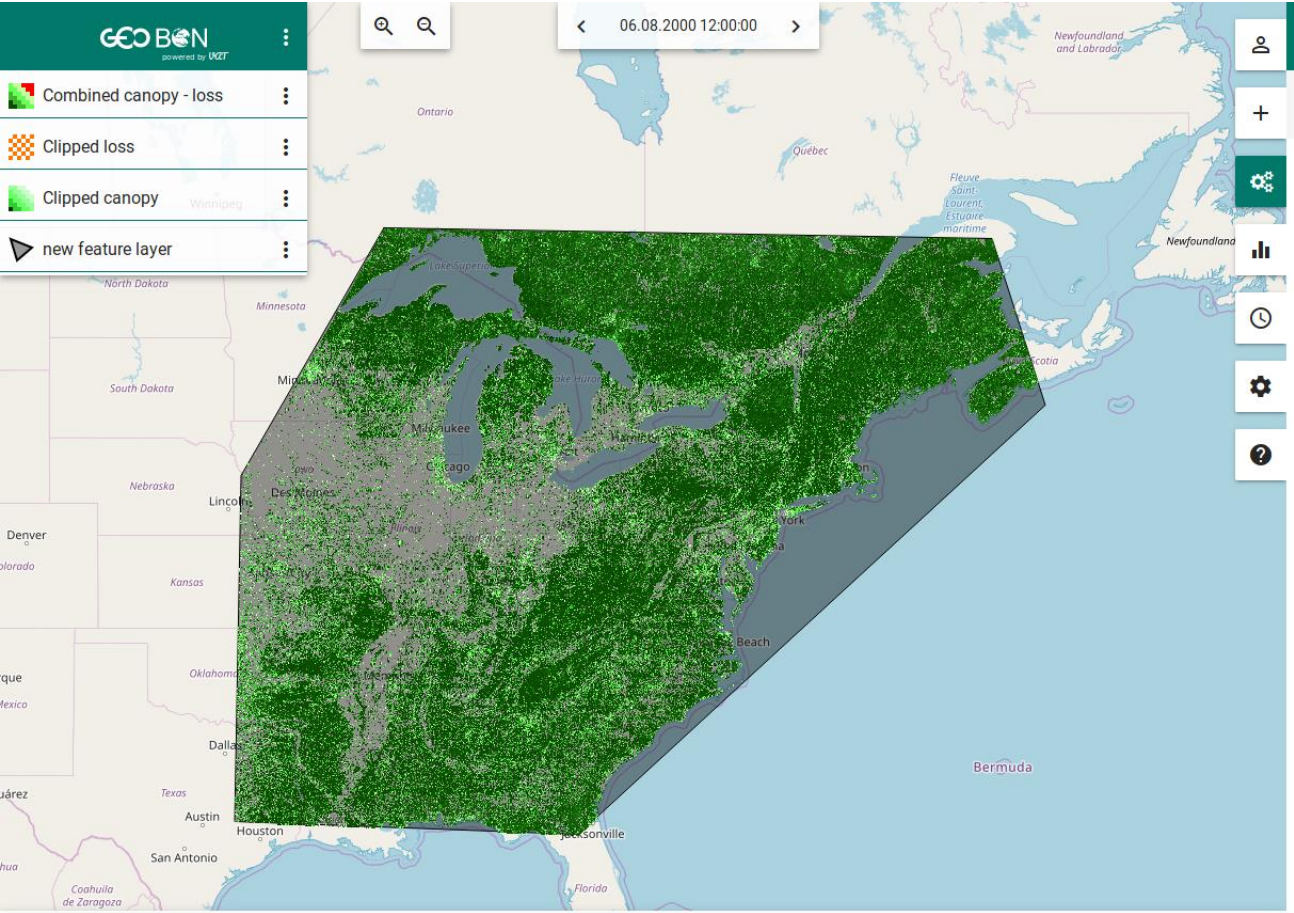
Maximum Value

254

Output Unit

loss (classification)

Create



Operators

Search

Mixed

- **Raster Polygon Clip** — Clip a raster image via polygon boundaries
- **Raster Value Extraction** — Attach raster values to vector data
- **R Script** — Execute an R script (experimental)

Plots

- **Box Plot** — Box plot your data
- **Histogram** — Create a histogram from vector or raster data
- **Pie Chart** — Plot your data as a pie chart
- **Scatter Plot** — Scatter plot your data

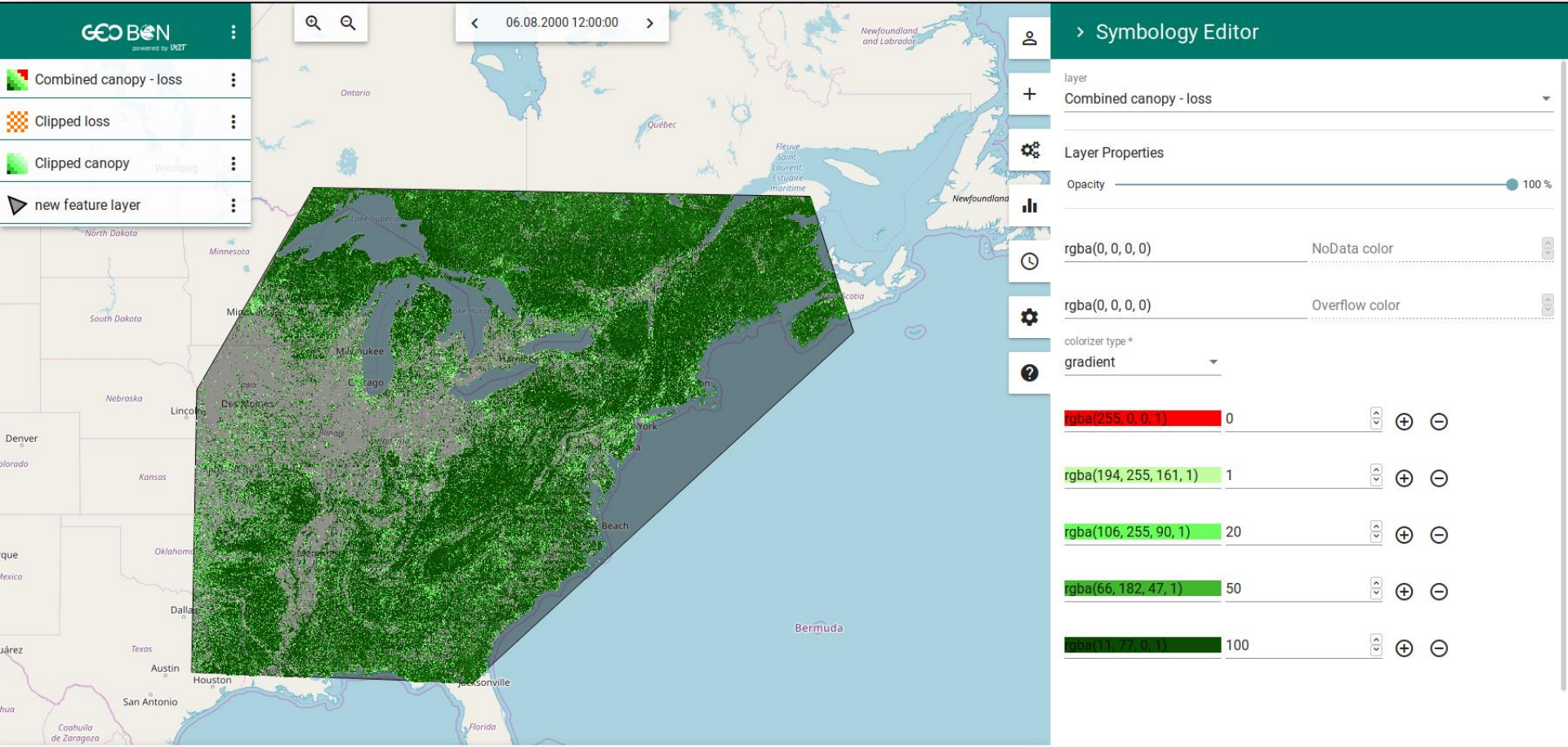
Raster

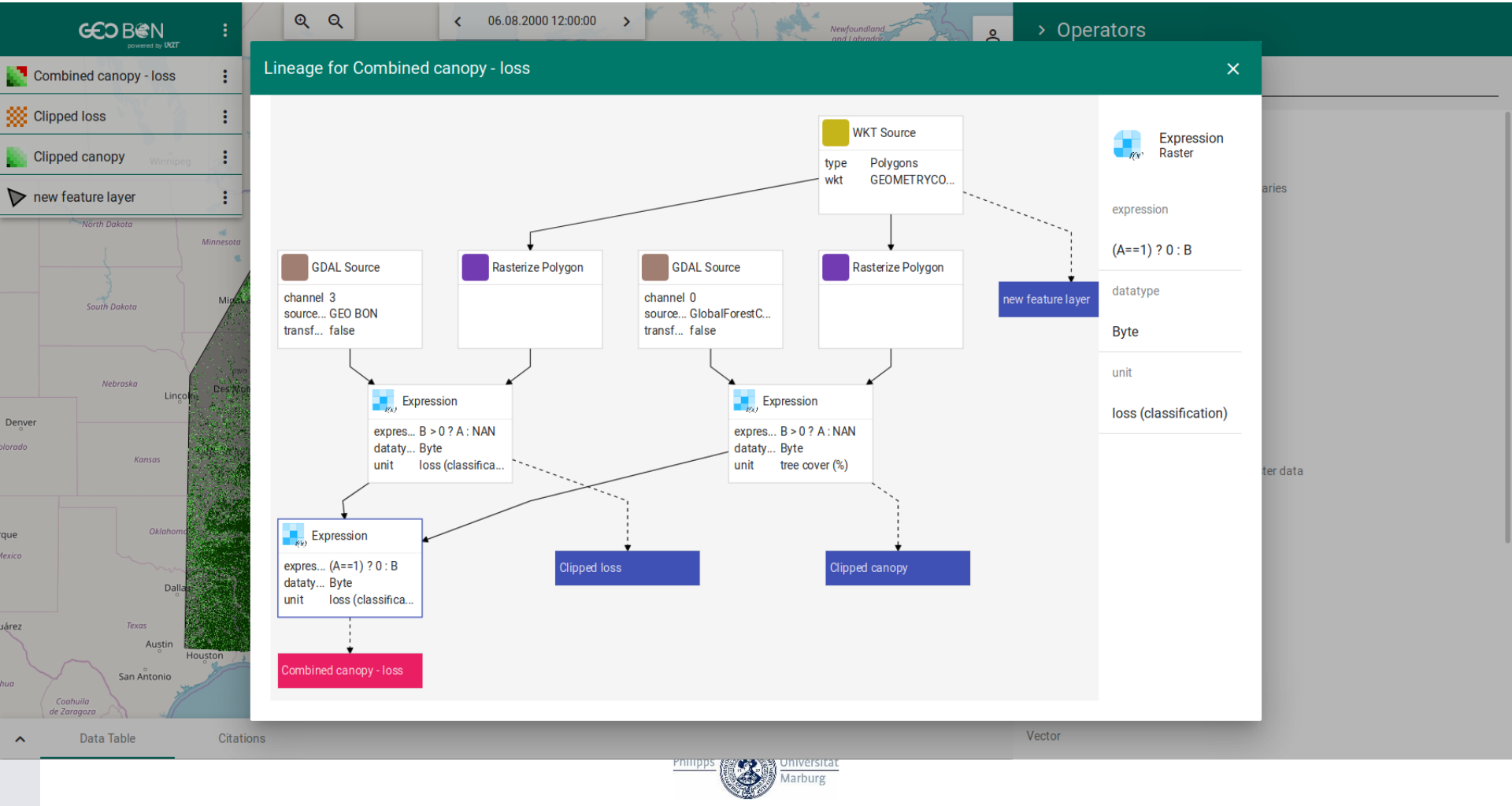
- **Expression** — Calculate an expression on a raster

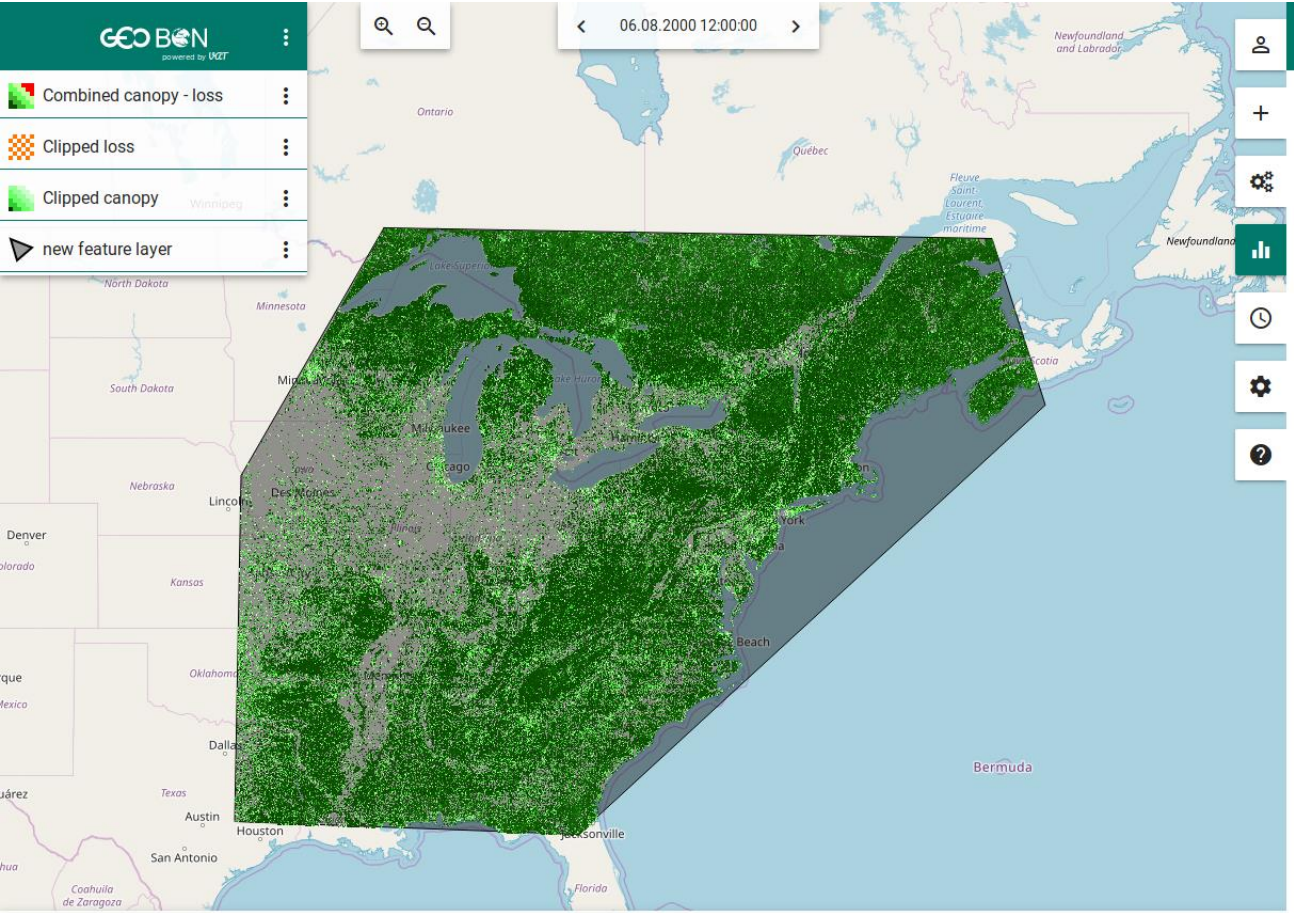
Vector

Data Table | Citations

Philipps Universität Marburg


GEO BON
powered by UZH
Combined canopy - loss
Clipped loss
Clipped canopy
new feature layer
06.08.2000 12:00:00
Symbology Editor
layer
Combined canopy - loss
Layer Properties
Opacity
100 %
rgba(0, 0, 0, 0)
NoData color
rgba(0, 0, 0, 0)
Overflow color
colorizer type *
gradient
rgba(255, 0, 0, 1)
0
rgba(194, 255, 161, 1)
1
rgba(106, 255, 90, 1)
20
rgba(66, 182, 47, 1)
50
rgba(11, 77, 0, 1)
100
Data Table
Citations
Ontario
Québec
Newfoundland and Labrador
Minnesota
North Dakota
South Dakota
Nebraska
Kansas
Oklahoma
Texas
Denver
Chicago
Milwaukee
Lincoln
Des Moines
Dallas
Austin
Houston
San Antonio
Jacksonville
Bermuda
Florida
Coahuila de Zaragoza

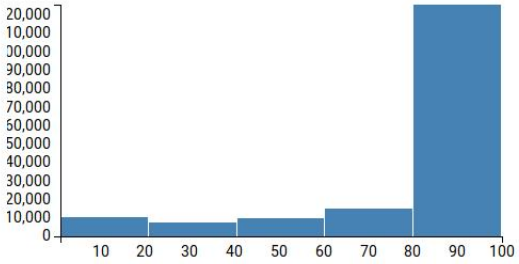

GEO BON
powered by UZH
Combined canopy - loss
Clipped loss
Clipped canopy
new feature layer
06.08.2000 12:00:00
Operators
Lineage for Combined canopy - loss
WKT Source
type Polygons
wkt GEOMETRYCO...
GDAL Source
channel 3
source... GEO BON
transf... false
Rasterize Polygon
GDAL Source
channel 0
source... GlobalForestC...
transf... false
Rasterize Polygon
new feature layer
Expression
expres... B > 0 ? A : NAN
dataty... Byte
unit loss (classifica...
Expression
expres... B > 0 ? A : NAN
dataty... Byte
unit tree cover (%)
Expression
expres... (A==1) ? 0 : B
dataty... Byte
unit loss (classifica...
Clipped loss
Clipped canopy
Combined canopy - loss
Expression
Raster
expression
(A==1) ? 0 : B
datatype
Byte
unit
loss (classification)
Data Table
Citations
Vector
Philipps Universität Marburg



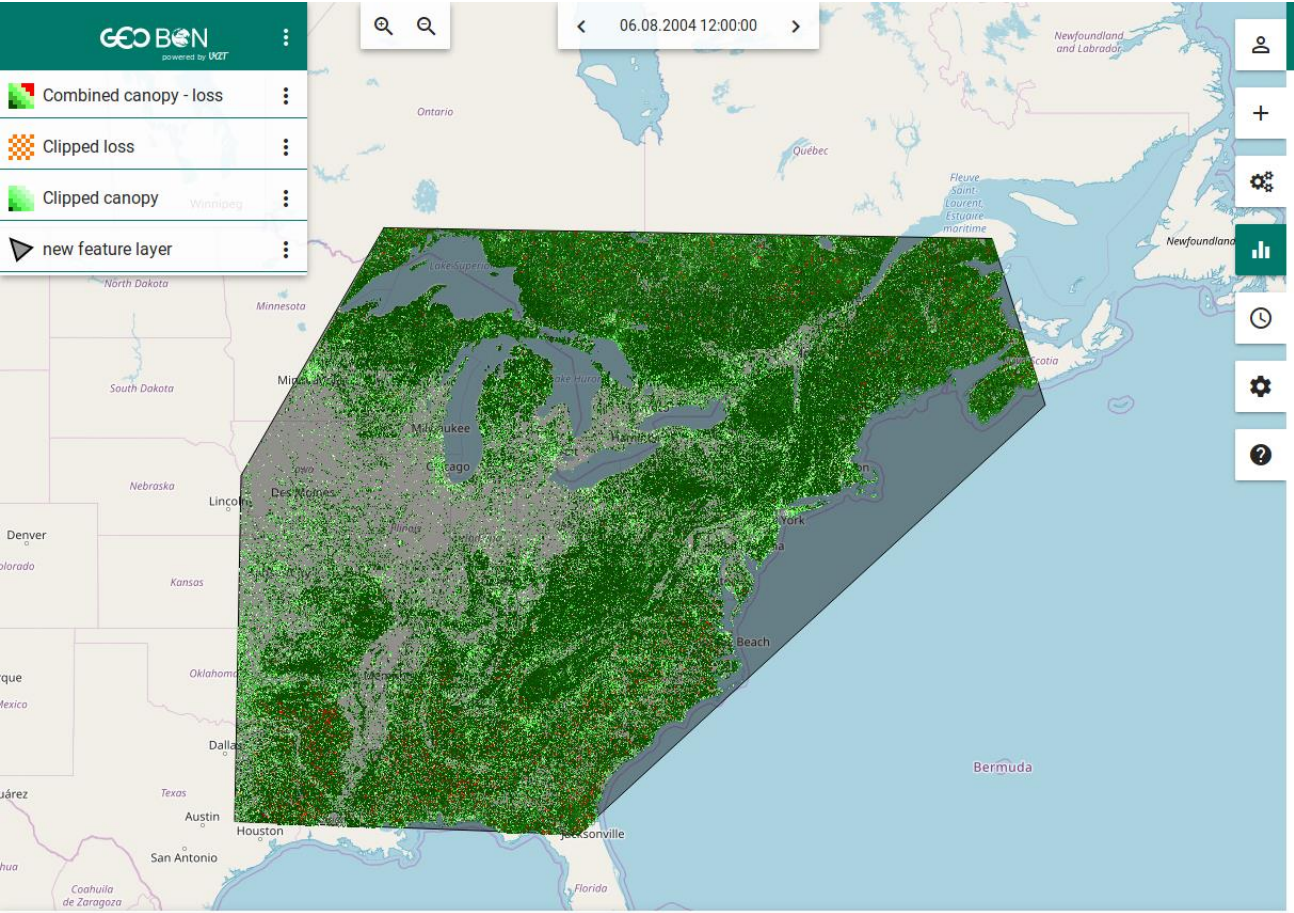
Plots

Histogram of Combined canopy - loss

Histogram

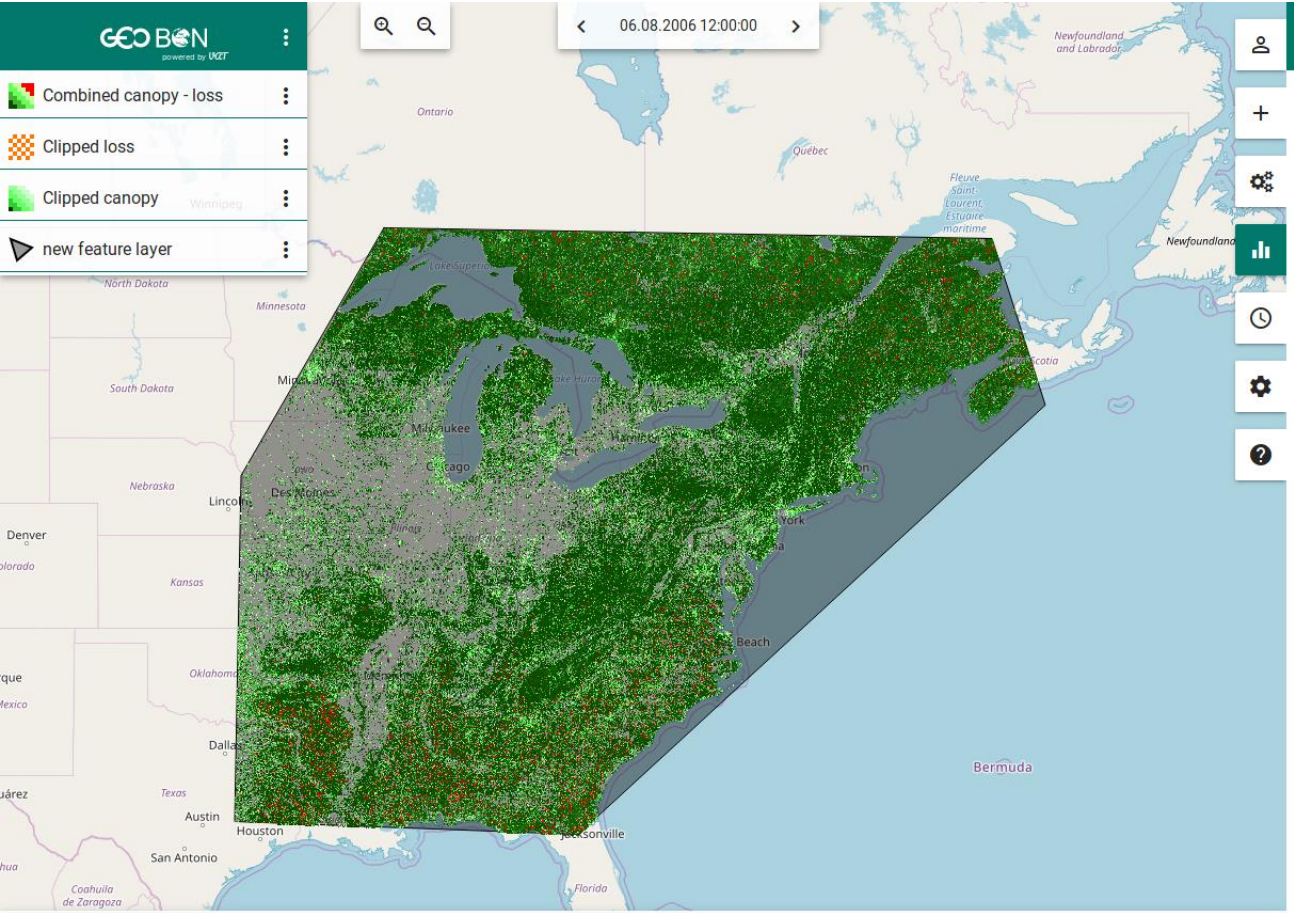


Philipps Universität Marburg



Data Table Citations

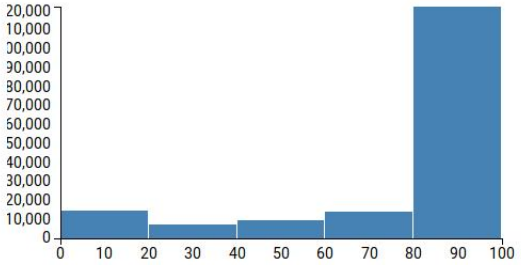
Philipps Universität Marburg



Plots

Histogram of Combined canopy - loss

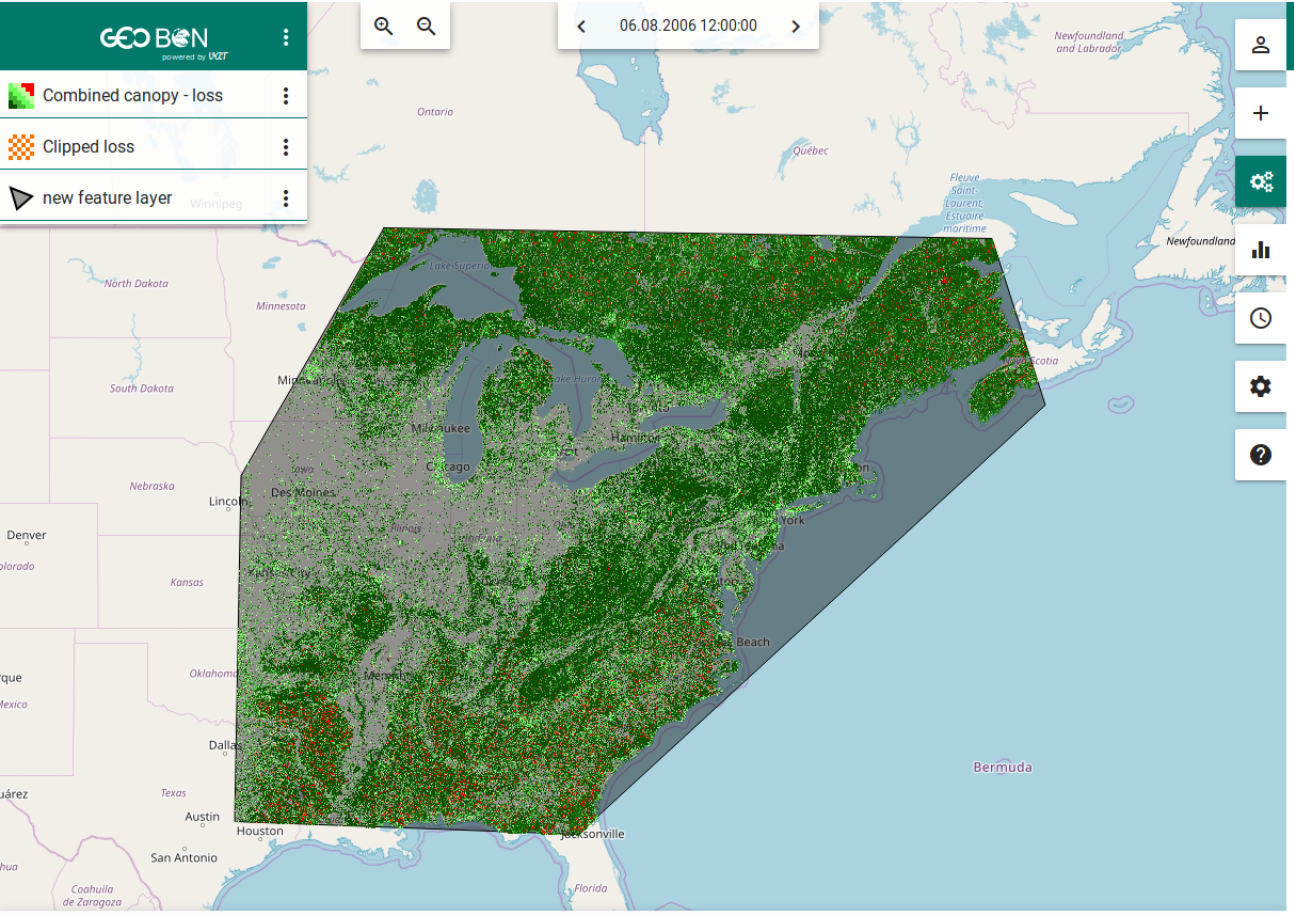
Histogram



Data Table

Citations

Philipps Universität Marburg



Data Table | Citations

Philipps Universität Marburg

R Operator

Raster

Select input Raster

Input A

Clipped loss

Configuration

Specify the operator

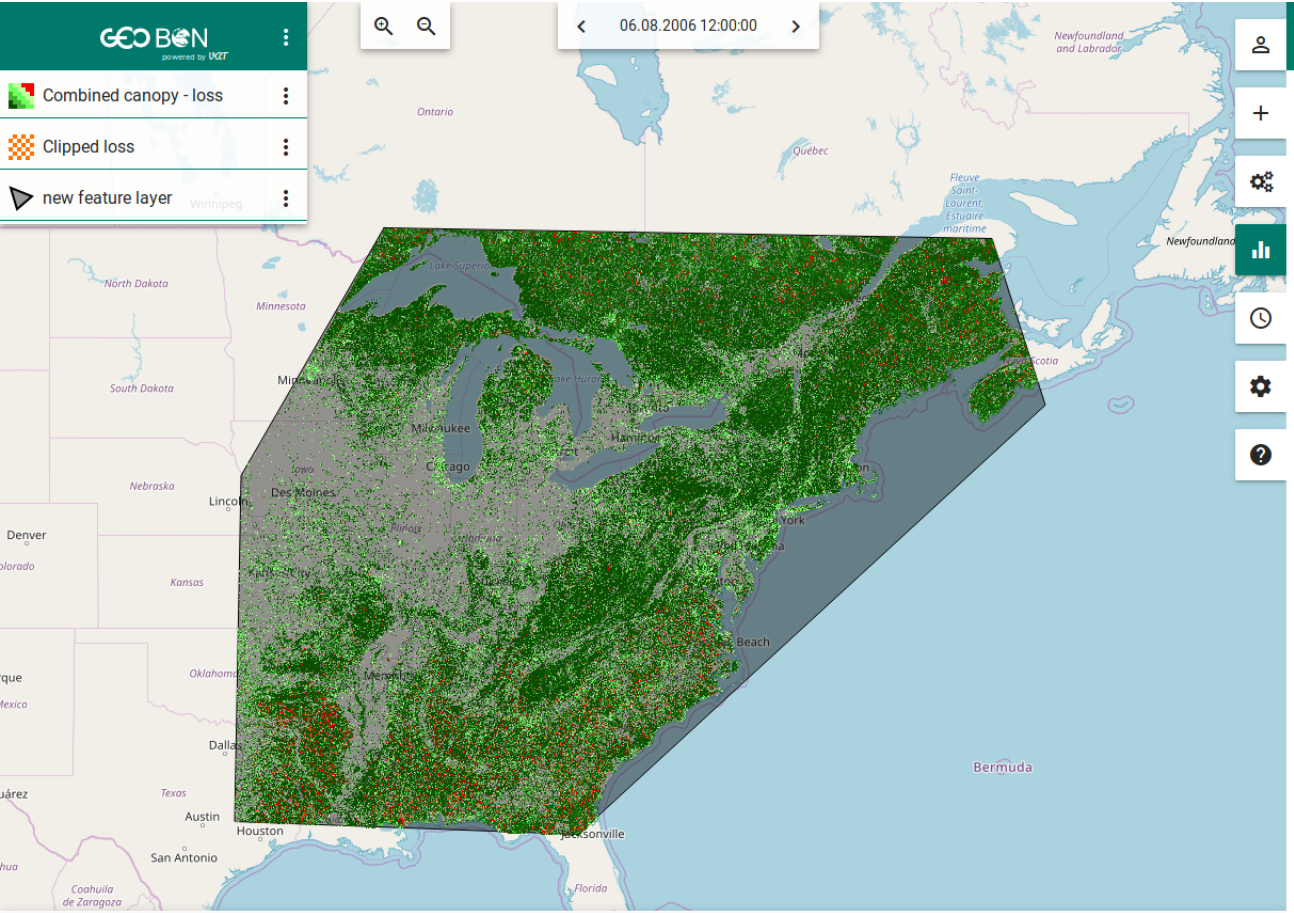
Quick Help — How the operator works

```
# Get the raster from Mapping / VAT backend, mapping.qrect will be replaced by the requested area in the UI!
raster_layer_1 <- mapping.loadRaster(0, mapping.qrect)

# Calculate the frequencys of each raster value (only usefull for classificaions)
raster_value_freqs <- freq(raster_layer_1)

print(raster_value_freqs)
# Calculate the fraction of cells with value == 1;
fraction_of_class <- raster_value_freqs[2,2] / (raster_value_freqs[1,2] + raster_value_freqs[2,2])
# Print the fraction ...
print(fraction of class)
```

Result Type

Text

Create



Plots

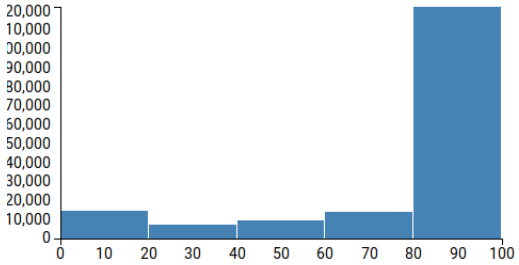
R Output

R Script

```
     value  count
[1,]     0 319841
[2,]     1   6323
[3,]   NaN 722412
     count
0.01938595
```

Histogram of Combined canopy - loss

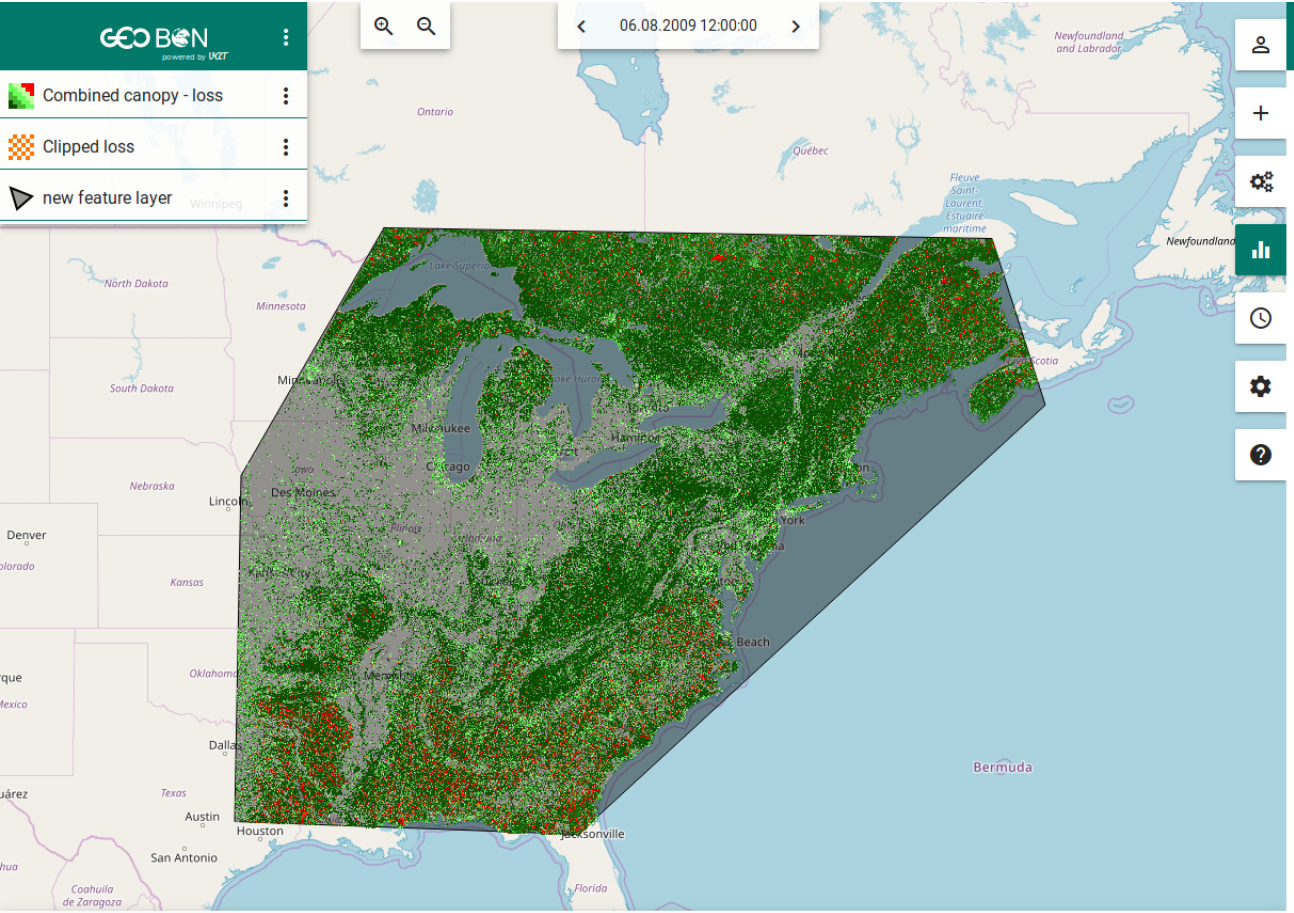
Histogram



Data Table

Citations

Philipps Universität Marburg



Plots

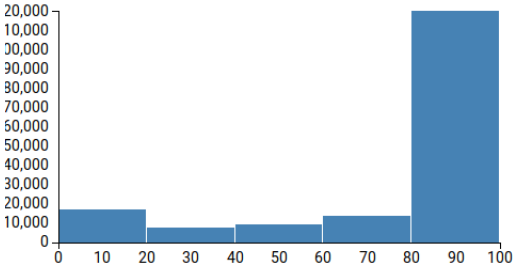
R Output

R Script

```
     value  count
[1,]     0 316628
[2,]     1   9536
[3,]   NaN 722412
     count
0.02923683
```

Histogram of Combined canopy - loss

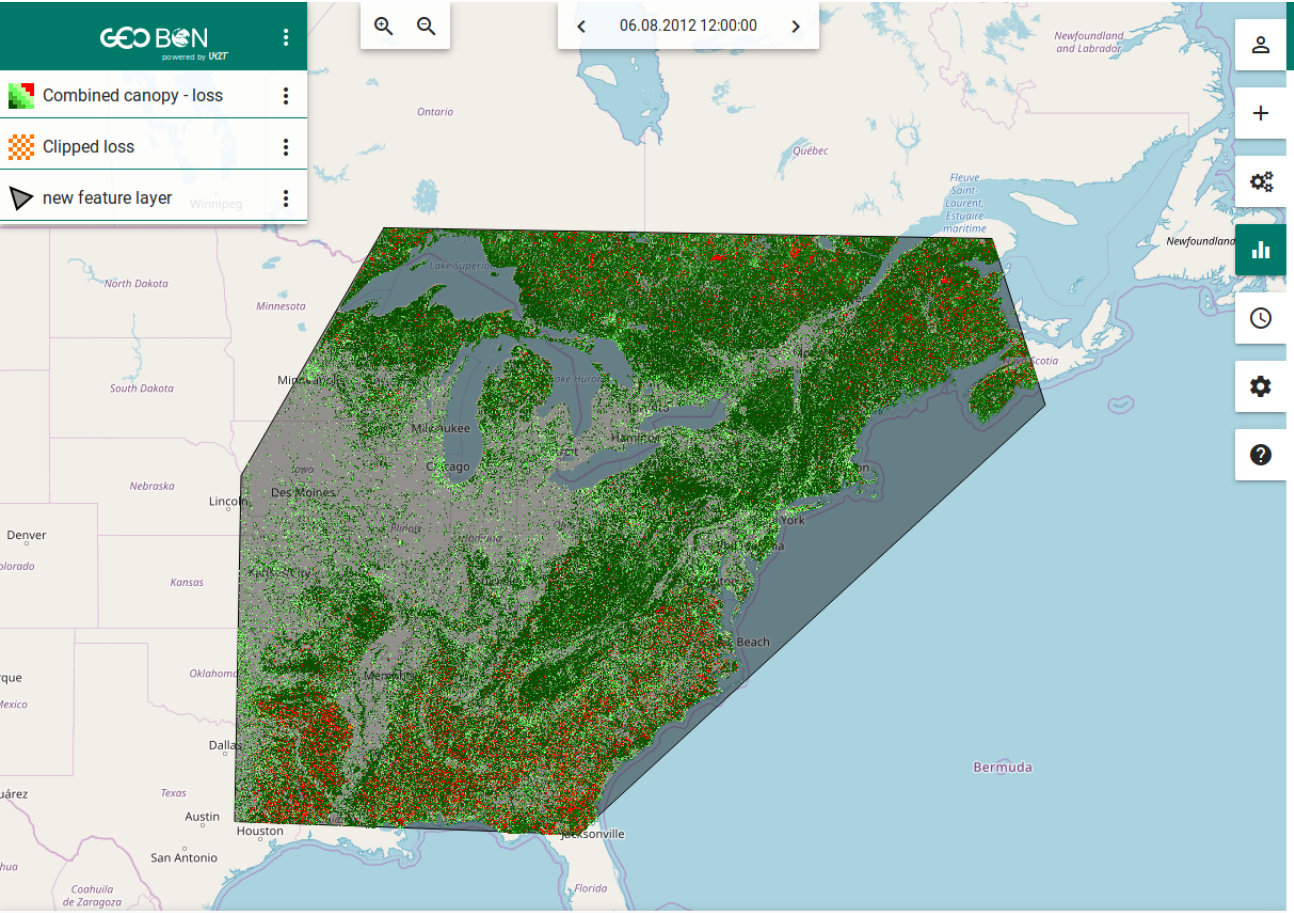
Histogram



Data Table

Citations

Philipps Universität Marburg



Plots

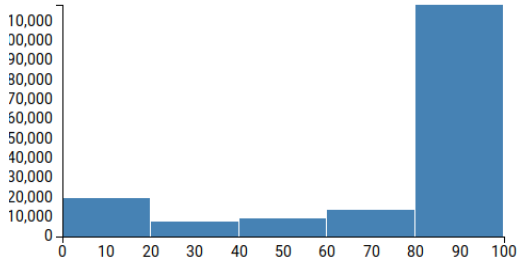
R Output

R Script

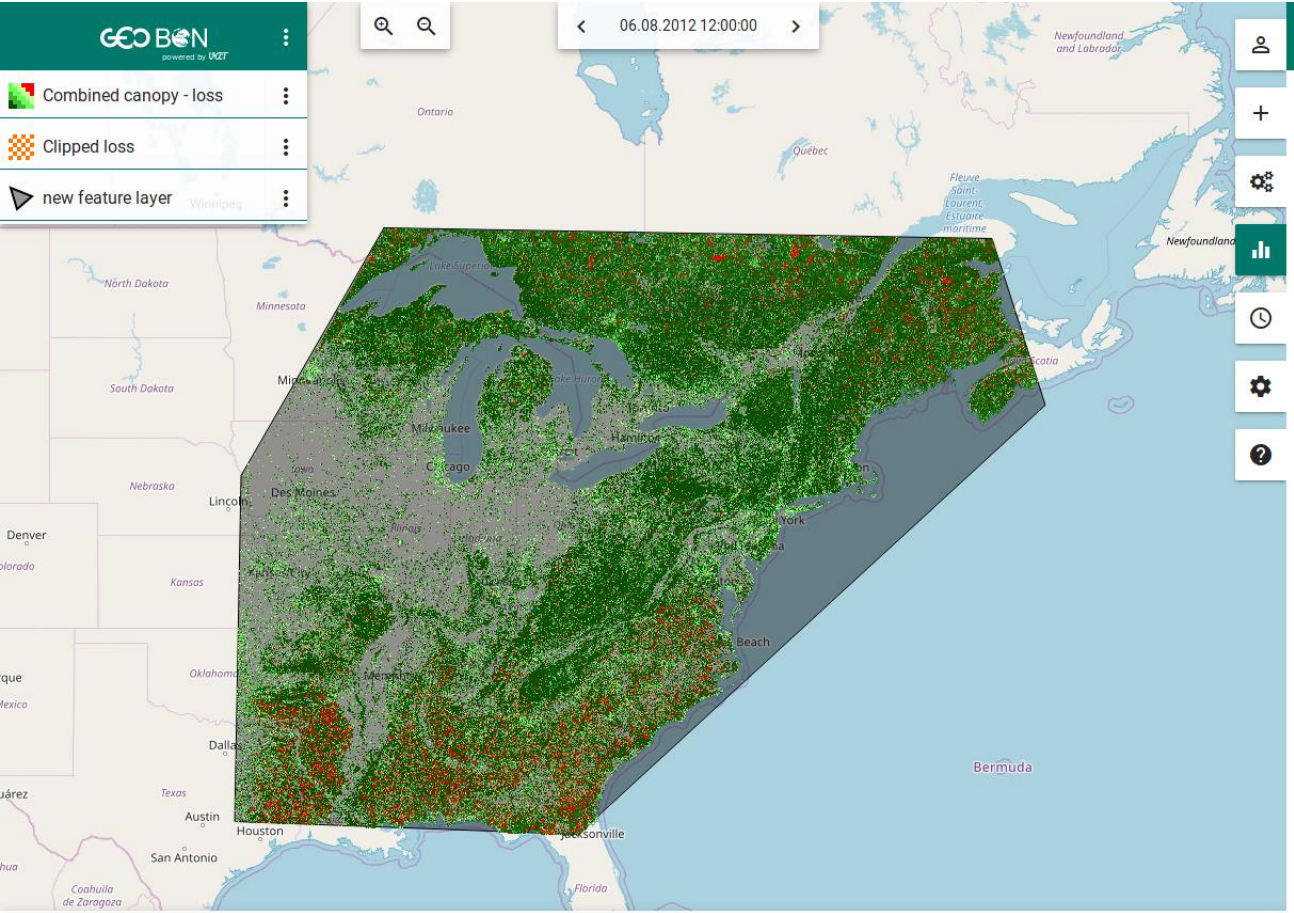
```
     value  count
[1,]     0 313564
[2,]     1  12600
[3,]   NaN 722412
     count
0.03863087
```

Histogram of Combined canopy - loss

Histogram

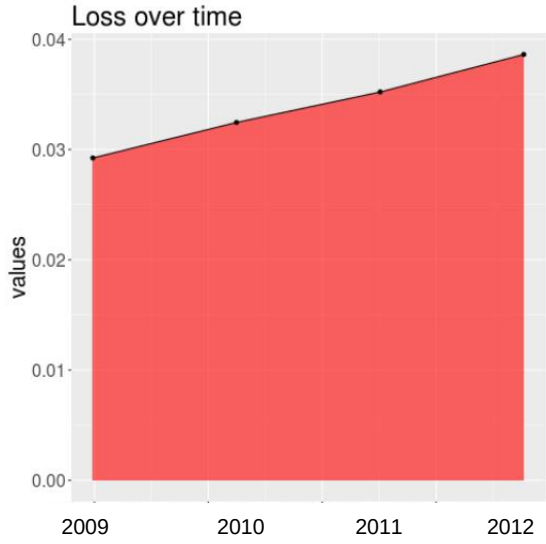


Philipps Universität Marburg



Plots

R Output
R Script



R Output
R Script

```
     value  count
[1,]     0 313564
[2,]     1  12600
```

Philipps Universität Marburg

Application Builder / Report Generator

- Workflows creation is a task for experts / researchers…

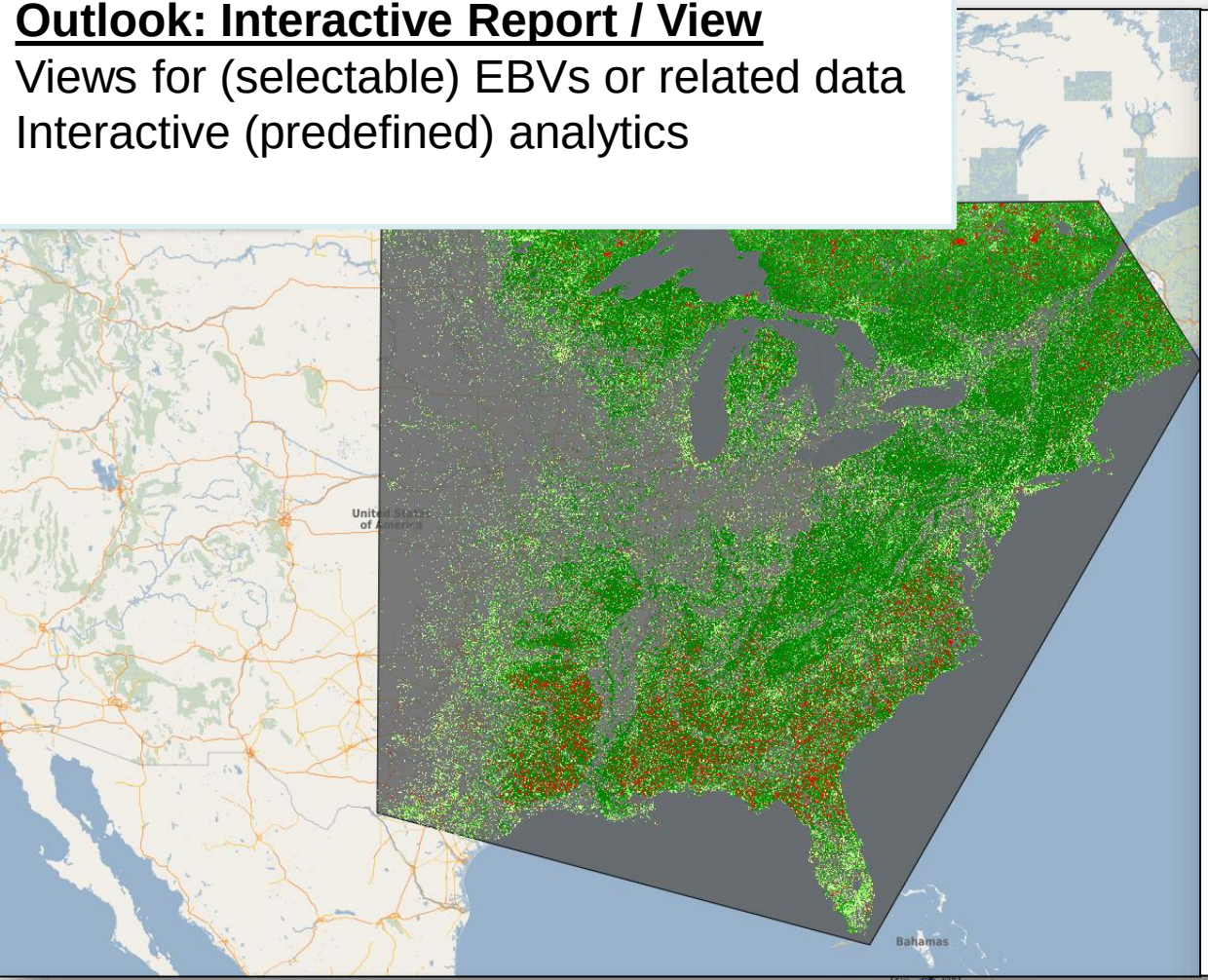
- How to …
  - display EBV information for end-users & decision makers?
  - make created analytics available in an intuitive way?

→ We will add an application builder / report generator:
Generate easy to use interactive visualizations!

Outlook: Interactive Report / View

Views for (selectable) EBVs or related data

Interactive (predefined) analytics



Select data layer

Global Forest Change loss

(Organization: University of Maryland; Published by Hansen, Potapov, Moore, Hancher et al.)

Draw or select area of interest

Feature type

Polygon

Start | End | Cancel

Year to display

2006

Analytics

Histogram of Tree cover - loss

Histogram

10,000
00,000
90,000
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0

0 10 20 30 40 50 60 70 80 90 100

Summary

- Domain experts explore data & create analytic workflows
  - Privacy preserving / combine EBVs easily with non-public data
  - Support of provenance & lineage (reproduce)
- Outlook: End-users & decision makers → Interactive reports / views!

- Reuse of existing functionality / avoid re-implementation

- Feedback and/or ideas?
  - vat@informatik.uni-marburg.de